Estimation and Robustness of Linear Mixed Models in Credibility Context

by Wing Kam Fung and Xiao Chen Xu

ABSTRACT

In this paper, linear mixed models are employed for estimation of structural parameters in credibility context. In particular, Hachemeister's model and Dannenburg's crossed classification model are considered. Maximum likelihood (ML) and restricted maximum likelihood (REML) methods are developed to estimate the variance and covariance parameters. These estimators are compared with the classical Hachemeister's and the Dannenburg's estimators by simulation. The robustness properties of the ML and REML methods are also investigated. In the simulation studies, we have tested the performance of ML, REML, and the classical estimation approaches when the error terms are normally distributed and lognormally distributed. It is noticed that the proposed ML and REML approaches have clear advantages over the classical estimation approaches. The mean-squared errors of the proposed estimators can be a few hundred times smaller than those of classical estimators.

1. Introduction

Credibility theory is a method to predict the future exposures of a risk entity based on past information. In statistics, the credibility data can be treated as longitudinal data, and the development of credibility theory has been closely linked to the longitudinal data model. Frees et al. (1999) has demonstrated the implementation of the linear mixed model under the classical credibility framework. The implementation of the generalized linear mixed model, which is an extension of the linear mixed model, has been proposed by Antonio and Beirlant (2006). Although only independent error structure has been considered in both literatures, the longitudinal data interpretation suggests additional techniques that actuaries can use in credibility rate making.

Later developments of credibility theory have considered the correlation between error terms. For instance, Cossette and Luong (2003) employed the regression credibility model, which can be regarded as a special form of the linear mixed model, to catch the random effects and within-panel correlation structure, and used weighted least squares method to estimate the variance covariance parameters. Lo, Fung, and Zhu (2006) and Lo, Fung, and Zhu (2007) proposed the generalized estimating equations (GEE) to handle the correlated error structure and estimate the variance of the random components under the regression credibility model. The methods in those papers have been justified by empirical studies.

In this paper, our attention is given to the linear mixed modeling in credibility context under Hachemeister's model and Dannenburg's model while taking into account both independent and correlated error structures. Maximum likelihood (ML) and restricted maximum likelihood (REML) methods are used to estimate the variance covariance parameters, where the random components are regarded as normally distributed. The performance of the ML and REML estimators are compared with the classical Hachemeister's and Dannenburg's estimators in simulation studies, when the error terms are normally distributed and non-normally distributed. In both situations, it can be shown that the ML that an REML approaches has clear advantages over its alternatives.

The structure of this paper is as follows. In Section 2, the regression credibility model is specified. Several commonly used error structures for modeling the observations of a risk entity are introduced. Section 3 gives a brief introduction on ML and REML methods, and their applications to the linear mixed model. The estimation of the structural parameters in Hachemeister's and Dannenburg's two-way crossed classification models are studied in Sections 4 and 5. In both sections, a brief introduction of the credibility model and classical estimation method is given, then two simulation studies are presented to examine the performance of the proposed ML and REML approaches. The first study tests the performance of ML and REML approaches when the observations are normally distributed. The second study tests the performance of ML and REML approaches when the observations are lognormally distributed, i.e., the normality assumption is violated. A few concluding remarks are given in the last section. It can be shown that enormous discrepancies of the performance of the credibility estimator for the credibility factors and future exposure between the classical estimation approach and the ML, REML approaches occur in both Dannenburg's model and Hachemeister's model. For instance, when the error terms follow multivariate normal distribution, the mean squared errors of the classical estimators for future exposure are a few hundred times higher than the counterpart in the proposed ML approach.

2. Model specification

2.1. Regression credibility model

In this paper, we employ the regression credibility model which is proposed by Hachemeister (1975). It is a specific form of a linear mixed model that can help us capture within-panel correlation. The regression model has the following form:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, 2, \ldots, n. \qquad (1)$$

Each element y_{ij} in the $r_i \times 1$ vector $\mathbf{y}_i$ corresponds to the observed value with regard to risk entity i in the jth observation period. The design matrix $\mathbf{X}_i$, of dimension $r_i \times m$, enters the model as a known constant matrix. The dimension of the vector of regression coefficients $\boldsymbol{\beta}_i$ is m. $\boldsymbol{\beta}_i$s are assumed to be independent and normally distributed, with common mean $\boldsymbol{\beta}$ and variance covariance matrix $\mathbf{F}$ for all i. The error vectors $\boldsymbol{\varepsilon}_i$s are taken to be independently distributed from a normal distribution with mean $\mathbf{0}$ and variance covariance matrix $\sigma^2 \mathbf{V}_i = \sigma^2 \mathbf{W}_i^{-1/2} \boldsymbol{\Gamma}_i \mathbf{W}_i^{-1/2}$, where $\mathbf{W}_i^{-1/2}$ is a diagonal weight matrix of known constants and $\boldsymbol{\Gamma}_i$ is a correlation matrix. Here we assume $\boldsymbol{\Gamma}_i$, which describes the correlation between the error terms ε_{ij}s for entity i, to be positive definite and depends on some fixed unknown parameters which are to be estimated. Aided by the specifications stated above, readers may easily derive the following about $\mathbf{y}_i$:

(a) $\mathbf{y}_i$ and $\mathbf{y}_j$ are statistically independent for $i \neq j$;
(b) $\boldsymbol{\mu}_i = E(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\beta}$;
(c) $\mathbf{V}(\mathbf{y}_i) = \mathbf{X}_i \mathbf{F} \mathbf{X}_i' + \sigma^2 \mathbf{W}_i^{-1/2} \boldsymbol{\Gamma}_i \mathbf{W}_i^{-1/2}$.

Hachemeister (1975) and Rao (1975) give the linear Bayes estimator for $\boldsymbol{\beta}_i$, which minimizes the mean-squared error losses. This estimator takes the following form:

$$\hat{\boldsymbol{\beta}}_i^{(B)} = \mathbf{Z}_i \hat{\boldsymbol{\beta}}_i^{(GLS)} + (\mathbf{I} - \mathbf{Z}_i) \boldsymbol{\beta}, \qquad (2)$$

where $\mathbf{Z}_i$ is the credibility matrix, $\hat{\boldsymbol{\beta}}_i^{(GLS)}$ is the generalized least squares estimator for $\boldsymbol{\beta}_i$, and we have

$$\mathbf{Z}_i = \mathbf{F}[\mathbf{F} + \sigma^2 (\mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}]^{-1}, \qquad (3)$$

$$\hat{\boldsymbol{\beta}}_i^{(GLS)} = (\mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i. \qquad (4)$$

As we can see from the above, in order to get the estimation of $\boldsymbol{\beta}_i$, we have to estimate the parameters σ^2, ρ, $\boldsymbol{\beta}$ and $\mathbf{V}_i$. The accuracy of the estimation of these parameters can largely affect the estimation efficiency for $\boldsymbol{\beta}_i$.

2.2. Several commonly used error structures

The moving average (MA), autoregressive (AR), and exchangeable types of error are commonly used to model the correlation structure of observations within a risk entity. Those structures have certain simplicity, and by using relatively few unknown parameters they can capture the correlation structure well. Therefore, under credibility frameworks, we could use all these correlation structures to model the correlation between error terms. However, in our empirical studies we would like to only incorporate the MA(1) and the exchangeable error correlation structures under each credibility framework for brevity.

2.2.1. Moving average correlation structure

For an MA(q) process, the correlation between the errors ε_j and ε_k can be written as

$$\Gamma_{jk} = \begin{cases} 1, & \text{for} \quad j = k, \\ \rho_{|j-k|}, & \text{for} \quad 0 < |j - k| \leq q, \\ 0, & \text{otherwise.} \end{cases}$$

For instance, the correlation matrix $(\Gamma_{jk})_{n \times n}$ of the MA(1) takes the explicit form of

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \ddots & \\ 0 & \rho & 1 & \ddots & \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & \rho & 1 \end{bmatrix}.$$

2.2.2. Autoregressive correlation structure

$AR(q)$ is given by the equation

$$\varepsilon_t = \sum_{i=1}^{q} \varphi_i \varepsilon_{t-i} + e_t.$$

As we can see there is no simple form for the correlation matrix when q gets large. Therefore $AR(1)$ is the most commonly used model. For the $AR(1)$ model, the correlation matrix $(\Gamma_{jk})_{n\times n}$ for the random errors ε_ts can be written in the following form:

$$\Gamma = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & & \ddots \\ \rho^2 & \rho & 1 & & \ddots \\ & \ddots & \ddots & \ddots & \\ \rho^{n-1} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}.$$

2.2.3. Exchangeable correlation structure

The exchangeable type of correlation is also known as the uniform correlation. The correlation matrix $(\Gamma_{jk})_{n\times n}$ of the exchangeable type of error can be written as:

$$\Gamma_{jk} = \begin{cases} 1, & \text{for} \quad j = k, \\ \rho, & \text{otherwise.} \end{cases}$$

Therefore the exchangeable correlation matrix takes the explicit form of

$$\Gamma = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & & \ddots \\ \rho & \rho & 1 & & \ddots \\ & \ddots & \ddots & \ddots & \\ \rho & \cdots & \rho & \rho & 1 \end{bmatrix}.$$

3. The ML and REML methods

In the regression credibility model, the variance and covariance parameters can be estimated using the well-known maximum likelihood (ML) and the restricted maximum likelihood (REML)

estimation methods. As we all know that maximum likelihood estimators are obtained by maximizing the likelihood function, the restricted maximum likelihood has been proposed by modifying the maximum likelihood by partitioning the likelihood under normality into two parts, one of which is free of fixed effects. The restricted maximum likelihood estimators can be obtained by maximizing that part. While preserving the good properties of the ML estimators, the REML estimators have an additional property, which is to reduce the analysis variance for many, if not all, balanced data. Because both ML and REML methods are common statistical methods, detailed introduction is omitted in this paper.

From our assumption, the error vectors, ε_i, and regression coefficient vectors, β_i, are normally distributed. This implies $\mathbf{y}_i$ follows a multivariate normal distribution with derivable mean and variance covariance matrix

$$\mathbf{y}_i \sim N(\mathbf{X}_i\beta, \ \mathbf{X}_i\mathbf{F}\mathbf{X}_i' + \sigma^2\mathbf{W}_i^{-1/2}\Gamma_i\mathbf{W}_i^{-1/2}),$$

where $\mathbf{X}_i\beta$ is the fixed effect component of the linear mixed model. Hence we can derive the log likelihood and the restricted log likelihood function of $\mathbf{y}_i$. They have been shown as

$$L_{\text{ML}} = c_1 - \frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{V}(\mathbf{y}_i)| - \frac{1}{2}\sum_{i=1}^{n}\mathbf{r}_i'\mathbf{V}(\mathbf{y}_i)\mathbf{r}_i, \quad (5)$$

$$L_{\text{REML}} = c_2 - \frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{V}(\mathbf{y}_i)|$$

$$- \frac{1}{2}\log\left(\sum_{i=1}^{n}|\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i|\right) - \frac{1}{2}\sum_{i=1}^{n}\mathbf{r}_i'\mathbf{V}(\mathbf{y}_i)\mathbf{r}_i,$$

$$(6)$$

where

$$\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i\left(\sum_{i=1}^{n}\mathbf{X}_i' \cdot \mathbf{V}^{-1}(\mathbf{y}_i) \cdot \mathbf{X}_i\right)^{-1}$$

$$\times \left(\sum_{i=1}^{n}\mathbf{X}_i' \cdot \mathbf{V}^{-1}(\mathbf{y}_i) \cdot \mathbf{y}_i\right),$$

and c_1, c_2 are appropriate constants.

We define the vector α which contains all of the parameters of interest. For example $\alpha = (\theta_{11}, \theta_{12}, \ldots, \theta_{mm}, \sigma^2, \rho)'$, where $\theta_{11}, \theta_{12}, \ldots, \theta_{mm}$ indicate the entries that specify the covariance matrix $\mathbf{F}$. We could solve α by maximizing the log likelihood function with regard to α or by solving the score function

$$\frac{\partial L_{\text{ML}}}{\partial \alpha} = 0$$

for the ML approach, and

$$\frac{\partial L_{\text{REML}}}{\partial \alpha} = 0$$

for the REML approach. More details about the derivation of the likelihood and restricted likelihood functions, fixed and random effects, estimates of the variance and covariance components can be found in Laird and Ware (1982), McCulloch (1997) and Verbeke and Molenberghs (2000).

Computationally there are various ways to obtain the ML and the REML estimators, such as the Newton-Raphson method and the simplex algorithm. Details of those methods can be found in Lindstrom and Bates (1988) and Nelder and Mead (1965). There are also many statistical packages available that can be used to perform such estimation, such as Matlab, R, S+ and SAS.

4. Parameter estimation in Hachemeister's model

4.1. Hachemeister's model and method

Hachemeister's model also known as the regression credibility model was proposed by Hachemeister (1975). It has the form

$$E[\mathbf{y}_i(\Theta)] = \mathbf{X}_i'\beta_i, \qquad i = 1, 2, \ldots, n, \qquad (7)$$

where Θ denotes the unobservable risk characteristic associated with each risk entity, and the dimension of β_i is m. We have

$$\text{Var}(\mathbf{y}_i \mid \Theta) = s^2(\Theta)\mathbf{W}_i^{-1}.$$

The credibility factor matrix stated in Hachemeister (1975) is

$$\mathbf{Z}_i = (\mathbf{F}\mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i + \sigma^2\mathbf{I})^{-1}\mathbf{F}\mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i. \qquad (8)$$

A weighted least squares estimate of β can be obtained by:

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}. \qquad (9)$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} \qquad (10)$$

are two large single unites, formed by the design matrices and the vectors of observations respectively, and

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & & & \mathbf{0} \\ & \mathbf{W}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{W}_n \end{bmatrix}, \qquad (11)$$

is constructed with individual exposure matrices as building blocks along the principal diagonal.

An unbiased estimator of σ^2 takes the form

$$\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\hat{\sigma}_i^2$$

$$= n^{-1}(n-m)^{-1}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\hat{\beta}_i)'\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i'\hat{\beta}_i),$$

$$(12)$$

where $\hat{\beta}_i$ is the weighted least square estimator for $\mathbf{b}(\theta_i)$. The estimator for the covariance matrix $\mathbf{F}$ is somewhat more complex. Define

$$\mathbf{G} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\sum_{i=1}^{n}(\mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i)(\hat{\beta}_i - \hat{\beta})(\hat{\beta}_i - \hat{\beta})',$$

$$(13)$$

$$\mathbf{\Pi} = \mathbf{I} - \sum_{i=1}^{n}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i)(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

$$\times (\mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i). \qquad (14)$$

The unbiased estimator for $\mathbf{F}$ is

$$\mathbf{C} = \mathbf{\Pi}^{-1}[\mathbf{G} - (n-1)(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\hat{\sigma}^2]. \quad (15)$$

Since $\mathbf{F}$ is symmetric, we can take our estimator as

$$\hat{\mathbf{F}} = (\mathbf{C} + \mathbf{C}')/2. \quad (16)$$

4.2. Empirical studies

To estimate the structural parameters in Hachemeister's model, we can use R, which is handy, user-friendly, and freely available from the Internet. The simulation results we show in this section are obtained by using the subroutine *lme* in R.

In this section, we use two approaches to estimate the structural parameters.

1. Hachemeister: The classical Hachemeister estimators are computed.
2. ML: The maximum likelihood estimation is used to compute the structural parameters. Two ML estimators are used in this paper. They are linked with the independent and exchangeable error structures and are denoted by ML-I and ML-EX respectively.

From the simulation results, the performance of the ML approach and the REML approach is quite close. None of them performs universally better than the other. Therefore, for the sake of brevity, we only show the results of the ML approach.

In this part, two studies have been considered. Study 1 allows us to compare the performances of the ML estimator and Hachemeister's estimator when the joint distribution of the observations in each contract is multivariate normal. The ML estimators are associated with different error structures, namely, independent and exchangeable error structure. Study 2 assesses the estimation efficiency of the ML estimator and Hachemeister's estimator when the joint distribution of the observations in each contract is not multivariate normal, but multivariate log-

normal. The number of replicates in each study is 500.

4.2.1. Study 1

In the simulation studies under Hachemeister's framework, the number of entities n is set to be 25, and the number of observations in contract i is set to be 5 for each entity. The parameter values are taken as follows:

$$\boldsymbol{\beta} = (20, 10)', \qquad \sigma^2 = 4^2, \qquad \theta_{11} = 3^2,$$

$$\theta_{12} = 4, \qquad \theta_{22} = 3^2.$$

Here θ_{11}, θ_{22} are the diagonal elements of $\mathbf{F}$, while θ_{12} is the off-diagonal element of $\mathbf{F}$. Each weighting element w_{ij} is generated from a Poisson distribution with its mean λ_i following a uniform distribution defined in the interval $(5, 100)$. The explanatory variable x_{ij1} is set to be 1, while x_{ij2} is simulated from the normal distribution with variance 5 and around the mean level which is uniformly selected from the interval $(-5, 5)$.

As for the simulation results, we show the bias and mean square error (MSE) of Hachemeister's and ML for β_{i1}, β_{i2}, Z_{i11}, Z_{i12}, Z_{i21}, Z_{i22}, θ_{11}, θ_{12}, θ_{22} and σ^2.

Table 1 is associated with an independent error structure, while Table 2 is associated with an MA(1) error structure. As we can see, while the unbiased property of Hachemeister's estimator for the variance and covariance parameters is reasonably well exhibited, the huge discrepancies of the performance of the credibility estimators for β_i and $\mathbf{Z}_i$ between the ML method and Hachemeister's method occur.

Notice that the mean squared error in estimating β_i is impressively low for the ML method under both the independent and the MA(1) error structures. Judging from the credibility formula for computing β_i, the accuracy of the estimation for β_i largely depends on the accuracy in estimating the credibility factor $\mathbf{Z}_i$. The estimation of $\mathbf{Z}_i$ relies on the estimation of the vari-

Table 1. Estimation results for Study 1 in the Hachemeister model associated with an independent error structure and the observation are simulated from normal distribution

Parameter		Method		
		ML-I	ML-MA1	Hachemeister
β_{i1}	Bias	-3.51×10^{-3}	-3.57×10^{-3}	1.73
	MSE	4.70×10^{-1} ($> 50000^{\dagger}$)	4.82×10^{-1} (> 50000)	4.26×10^{4}
β_{i2}	Bias	-2.16×10^{-3}	-1.87×10^{-3}	-1.83×10^{-1}
	MSE	5.34×10^{-2} (> 10000)	5.46×10^{-2} (> 10000)	6.05×10^{2}
Z_{i11}	Bias	-1.12×10^{-2}	-9.16×10^{-3}	-5.32×10^{-1}
	MSE	1.16×10^{-3} (> 1000000)	1.17×10^{-3} (> 1000000)	3.93×10^{3}
Z_{i12}	Bias	4.52×10^{-3}	3.63×10^{-3}	3.41×10^{-1}
	MSE	6.85×10^{-4} (> 1000000)	6.70×10^{-4} (> 1000000)	1.72×10^{3}
Z_{i21}	Bias	6.75×10^{-4}	5.53×10^{-4}	5.79×10^{-2}
	MSE	9.75×10^{-5} (> 500000)	1.00×10^{-4} (> 500000)	5.49×10^{1}
Z_{i22}	Bias	-1.22×10^{-3}	-9.88×10^{-4}	-3.62×10^{-2}
	MSE	7.57×10^{-5} (> 100000)	7.35×10^{-5} (> 100000)	2.42×10^{1}
θ_{11}	Bias	-4.37×10^{-1}	-4.37×10^{-1}	3.04×10^{-1}
	MSE	7.85 (11.3)	7.84 (11.3)	8.84×10^{1}
θ_{12}	Bias	-2.50×10^{-1}	-2.51×10^{-1}	2.59×10^{-1}
	MSE	3.84 (9.77)	3.83 (9.79)	3.75×10^{1}
θ_{22}	Bias	-4.89×10^{-1}	-4.91×10^{-1}	-2.04×10^{-1}
	MSE	6.44 (2.02)	6.43 (2.02)	1.30×10^{1}
σ^2	Bias	6.19×10^{-2}	2.85×10^{-2}	5.98×10^{-2}
	MSE	6.37 (1.00)	8.37 (0.76)	6.40

† Relative efficiency of the estimator. Hachemeister's estimator serves as the base line.

ance and covariance parameters. Thus, the mean squared error (MSE) for each of the parameters specifying **F** in Hachemeister's approach is two to eleven times higher than its counterpart in the ML estimation approach. From our simulation results, around 15% of the estimates of the covariance matrix **F** are found not to be positive. In contrast, the ML approach gives reasonable estimates for all structural parameters. The poor estimation of $\mathbf{Z}_i$ in Hachemeister's method is likely to be incurred by the low accuracy level in estimating the variance and covariance parameters. As a result, the huge squared error loss for $\boldsymbol{\beta}_i$ occurs.

From Table 1, we can see that the ML-I method has slight advantages to ML-MA1 method due to its correct assumption about the error structure, and the reverse is true for Table 2. Comparing to the classical method, the MSE of θ_{11},

θ_{12} and θ_{22} are reduced by 50% to 90% in the ML approach. This impressive improvement results in enormous reductions of MSE in estimating the credibility factors (relative efficiency beyond 500,000 in Table 1, relative efficiency beyond 5,000 in Table 2). Hence the estimation accuracy of $\boldsymbol{\beta}_i$ has been largely improved (relative efficiency beyond 10,000 in Table 1, relative efficiency beyond 450 in Table 2).

4.2.2. Study 2

In this study, while taking the same setting used in Study 1, the vectors of error terms are simulated from multivariate lognormal distribution. This distribution has skewness of 0.33 and kurtosis of 6.64. Therefore the simulation results in this study show us the performance of the proposed ML and Hachemeister's estimators when the observations are no longer normally distributed.

Table 2. Estimation results for Study 1 in the Hachemeister model associated with a MA(1) error structure ($\rho = 0.4$) and the observation are simulated from normal distribution

Parameter		Method		
		ML-I	ML-MA1	Hachemeister
β_{i1}	Bias	-7.81×10^{-3}	-9.48×10^{-3}	-0.159
	MSE	4.72×10^{-1} (463)	4.25×10^{-1} (514)	2.18×10^{2}
β_{i2}	Bias	-6.30×10^{-4}	4.93×10^{-4}	2.04×10^{-2}
	MSE	4.45×10^{-2} (948)	3.55×10^{-2} (> 1000)	42.2
Z_{i11}	Bias	-8.19×10^{-3}	-8.31×10^{-3}	-1.25×10^{-2}
	MSE	1.40×10^{-3} (> 5000)	1.00×10^{-3} (> 5000)	7.35
Z_{i12}	Bias	5.95×10^{-3}	3.44×10^{-3}	1.22×10^{-2}
	MSE	8.04×10^{-4} (> 5000)	5.39×10^{-4} (> 10000)	5.87
Z_{i21}	Bias	3.83×10^{-3}	5.67×10^{-5}	7.30×10^{-3}
	MSE	1.72×10^{-4} (> 5000)	5.51×10^{-5} (> 10000)	1.01
Z_{i22}	Bias	-3.64×10^{-3}	-4.93×10^{-4}	-1.66×10^{-2}
	MSE	1.32×10^{-4} (> 5000)	3.82×10^{-5} (> 10000)	1.17
θ_{11}	Bias	-3.56×10^{-1}	-4.01×10^{-1}	3.51×10^{-1}
	MSE	7.66 (11.0)	7.55 (11.2)	8.43×10^{1}
θ_{12}	Bias	-2.06×10^{-1}	-2.12×10^{-1}	2.51×10^{-1}
	MSE	3.93 (9.41)	3.94 (9.39)	3.70×10^{1}
θ_{22}	Bias	-5.09×10^{-1}	-5.09×10^{-1}	-2.10×10^{-1}
	MSE	6.39 (2.02)	6.38 (2.02)	1.29×10^{1}
σ^2	Bias	-2.46	5.47×10^{-2}	-2.46
	MSE	11.9 (1.00)	9.71 (1.23)	1.19×10^{1}

From Table 3, we can see the MSE of the structural parameters θ_{11}, θ_{12}, and θ_{22} in the ML approach have been reduced by 50% to 80% relative to Hachemeister's approach. There are enormous discrepancies in the performance of the estimation of the credibility factors between the ML approach and Hachemeister's approach. The relative efficiency is more than 100,000 for Z_{i11}, Z_{i12}, Z_{i21}, Z_{i22}. Hence the estimation of β_i has been largely improved in the ML approach. With reference to Table 4, the ML-MA1 method performs the best in estimating β_i and the credibility factors due to its correct assumption about the error structure. The relative efficiency for the credibility factors reaches the level beyond 1000, while the MSE of β_{i1} and β_{i2} in Hachemeister's approach is 26–60 times higher than the counterparts in the ML approach. Hence, we can see that though distribution of the error terms violate the assumptions made in the ML approach, they still perform very well compared to Hachemeister's approach.

5. Parameter estimation in the crossed classification model

5.1. Dannenburg's credibility model and method

Dannenburg et al. (1996) proposed the two-way crossed classification model. In Dannenburg's model, the risk factors are treated in a symmetrical way. The two-way crossed classification model takes the following form:

$$y_{ijt} = \beta + \alpha_i^{(1)} + \alpha_j^{(2)} + \alpha_{ij}^{(12)} + \epsilon_{ijt},$$
$$t = 1, \ldots, T_{ij}. \qquad (17)$$

In this model, there are two risk factors. The number of categories of the first factor is I and of the second risk factor is J. An insurance portfolio which is subdivided by these two risk factors can be viewed as a two-way table.

Table 3. Estimation results for Study 2 in the Hachemeister model associated with an independent error structure and the observation are simulated from lognormal distribution

Parameter		Method		
		ML-I	ML-MA1	Hachemeister
β_{i1}	Bias	1.25×10^{-3}	1.51×10^{-3}	3.88×10^{-1}
	MSE	3.12×10^{-1} (> 5000)	3.19×10^{-1} (> 5000)	1.96×10^{3}
β_{i2}	Bias	9.90×10^{-4}	1.26×10^{-3}	-7.49×10^{-2}
	MSE	3.01×10^{-2} (> 1000)	3.07×10^{-2} (> 1000)	5.11×10^{1}
Z_{i11}	Bias	-7.54×10^{-3}	-6.87×10^{-3}	-6.27×10^{-2}
	MSE	6.30×10^{-4} (> 100000)	7.04×10^{-4} (> 100000)	2.60×10^{2}
Z_{i12}	Bias	2.85×10^{-3}	2.66×10^{-3}	9.49×10^{-2}
	MSE	3.09×10^{-4} (> 1000000)	3.18×10^{-4} (> 1000000)	3.32×10^{2}
Z_{i21}	Bias	2.26×10^{-4}	2.29×10^{-4}	2.56×10^{-2}
	MSE	3.29×10^{-5} (> 100000)	3.34×10^{-5} (> 100000)	7.01
Z_{i22}	Bias	-6.93×10^{-4}	-6.13×10^{-4}	-3.30×10^{-2}
	MSE	1.88×10^{-5} (> 100000)	1.86×10^{-5} (> 100000)	8.61
θ_{11}	Bias	-4.20×10^{-1}	-4.28×10^{-1}	-5.81×10^{-3}
	MSE	7.29 (5.60)	7.29 (5.60)	4.08×10^{1}
θ_{12}	Bias	-2.22×10^{-1}	-2.23×10^{-1}	3.52×10^{-2}
	MSE	3.75 (5.79)	3.74 (5.80)	2.17×10^{1}
θ_{22}	Bias	-0.46×10^{-1}	-4.61×10^{-1}	-4.21×10^{-2}
	MSE	6.27 (2.19)	6.26 (2.19)	1.37×10^{1}
σ^{2}	Bias	5.89×10^{-3}	6.72×10^{-2}	2.11×10^{-3}
	MSE	7.11 (0.99)	8.74 (0.81)	7.06

Suppose I is 2, J is 3. We have

$$
\begin{array}{c}
\alpha_1^{(1)} + \rightarrow \quad \boxed{+\alpha_{11}^{(12)} \quad +\alpha_{12}^{(12)} \quad +\alpha_{13}^{(12)}} \\
\alpha_2^{(1)} + \rightarrow \quad \boxed{+\alpha_{21}^{(12)} \quad +\alpha_{22}^{(12)} \quad +\alpha_{23}^{(12)}} \\
\uparrow \qquad \uparrow \qquad \uparrow \\
+\alpha_1^{(2)} \quad +\alpha_2^{(2)} \quad +\alpha_3^{(2)}
\end{array}
$$

The first risk factor $\alpha_i^{(1)}$ can be called the row factor. The second risk factor $\alpha_j^{(2)}$ can be called the column factor. The structural parameters are defined as follows:

$$
\mathrm{Var}(\alpha_i^{(1)}) = b^{(1)}, \qquad \mathrm{Var}(\alpha_j^{(2)}) = b^{(2)},
$$

$$
\mathrm{Var}(\alpha_{ij}^{(12)}) = a, \qquad \mathrm{Var}(\epsilon_{ijt}) = s^2/w_{ijt}.
$$

The credibility estimator of $y_{ij,T_{ij}+1}$ is equal to (Dannenburg et al., 1996):

$$
y_{ij,T_{ij}+1} = \beta + z_{ij}(y_{ijw} - \beta) + (1 - z_{ij})z_i^{(1)}(x_{izw} - \beta)
$$

$$
+ (1 - z_{ij})z_j^{(2)}(x_{zjw} - \beta), \qquad (18)
$$

where the credibility factors are

$$
z_{ij} = \frac{a}{a + \sigma^2/w_{ij\Sigma}}, \qquad \text{with} \quad w_{ij\Sigma} = \sum_t w_{ijt}, \qquad (19)
$$

$$
z_i^{(1)} = \frac{b^{(1)}}{b^{(1)} + a/z_{i\Sigma}}, \qquad \text{with} \quad z_{i\Sigma} = \sum_j z_{ij}, \qquad (20)
$$

$$
z_j^{(2)} = \frac{b^{(2)}}{b^{(2)} + a/z_{\Sigma j}}, \qquad \text{with} \quad z_{\Sigma j} = \sum_i z_{ij}. \qquad (21)
$$

x_{izw}, x_{zjw} are the adjusted weighted averages, which can give us a much clearer view on the risk experience with regard to different risk factors,

$$
x_{izw} = \sum_j \frac{z_{ij}}{z_{i\Sigma}}(y_{ijw} - \Xi_j^{(2)*}), \qquad (22)
$$

$$
x_{zjw} = \sum_i \frac{z_{ij}}{z_{\Sigma j}}(y_{ijw} - \Xi_i^{(1)*}), \qquad (23)
$$

Table 4. Estimation results for Study 2 in the Hachemeister model associated with a MA(1) error structure ($\rho = 0.4$) and the observation are simulated from lognormal distribution

Parameter		Method		
		ML-I	ML-MA1	Hachemeister
β_{i1}	Bias	3.57×10^{-4}	-1.92×10^{-3}	1.01×10^{-3}
	MSE	3.47×10^{-1} (54.2)	3.12×10^{-1} (60.3)	1.88×10^{1}
β_{i2}	Bias	-4.75×10^{-4}	-1.31×10^{-3}	2.31×10^{-3}
	MSE	2.90×10^{-2} (26.6)	2.51×10^{-2} (30.7)	7.70×10^{-1}
Z_{i11}	Bias	-5.43×10^{-4}	-6.57×10^{-3}	-6.86×10^{-3}
	MSE	6.64×10^{-4} (> 1000)	6.12×10^{-4} (> 1000)	2.00
Z_{i12}	Bias	3.26×10^{-4}	2.35×10^{-3}	6.42×10^{-3}
	MSE	3.03×10^{-4} (> 1000)	2.87×10^{-4} (> 1000)	6.15×10^{-1}
Z_{i21}	Bias	9.17×10^{-4}	1.53×10^{-4}	1.49×10^{-3}
	MSE	4.26×10^{-5} (> 1000)	2.11×10^{-5} (> 1000)	6.84×10^{-2}
Z_{i22}	Bias	-1.07×10^{-3}	-4.62×10^{-4}	-3.74×10^{-4}
	MSE	1.95×10^{-5} (> 1000)	1.24×10^{-5} (> 1000)	2.40×10^{-2}
θ_{11}	Bias	-3.20×10^{-1}	-3.95×10^{-1}	1.10×10^{-1}
	MSE	7.57 (5.18)	7.60 (5.16)	3.92×10^{1}
θ_{12}	Bias	-1.98×10^{-1}	-2.20×10^{-1}	2.89×10^{-2}
	MSE	3.95 (5.57)	4.04 (5.45)	2.20×10^{1}
θ_{22}	Bias	-4.55×10^{-1}	-4.62×10^{-2}	-2.10×10^{-1}
	MSE	6.27 (2.19)	6.37 (2.15)	1.37×10^{1}
σ^2	Bias	-2.70	-1.17×10^{-1}	-2.70
	MSE	1.25×10^{1} (0.99)	8.89 (1.39)	1.24×10^{1}

where

$$y_{ijw} = \sum_t \frac{w_{ijt}}{w_{ij\Sigma}} y_{ijt}.$$

And $\Xi_i^{(1)*}$, $\Xi_j^{(2)*}$ are the row effect and the column effect respectively. They can be found as the solution of the following $I + J$ linear equations using iterative approach.

$$\Xi_i^{(1)*} = z_i^{(1)} \left[\sum_j \frac{z_{ij}}{z_{i\Sigma}}(y_{ijw} - \Xi_j^{(2)*}) - \beta \right], \tag{24}$$

$$\Xi_j^{(2)*} = z_j^{(2)} \left[\sum_i \frac{z_{ij}}{z_{\Sigma j}}(y_{ijw} - \Xi_i^{(1)*}) - \beta \right]. \tag{25}$$

In Dannenburg's approach, the structural parameters β and s^2 can be estimated by the following equations (Dannenburg et al.,

1996):

$$\beta = x_{www} = \sum_i \sum_j \frac{w_{ij\Sigma}}{w_{\Sigma\Sigma\Sigma}} y_{ijw}, \tag{26}$$

$$s^{2\bullet} = \frac{\sum_i \sum_j \sum_t w_{ijt}(y_{ijt} - y_{ijw})^2}{\sum_i \sum_j (T_{ij} - 1)_+}. \tag{27}$$

To obtain the estimators a, $b^{(1)}$ and $b^{(2)}$, Dannenburg et al. (1996) suggested to solve the following linear equations on moments:

$$E\left[\frac{1}{I} \sum_i \left(\sum_j \frac{w_{ij\Sigma}}{w_{i\Sigma\Sigma}}(y_{ijw} - y_{iww})^2 - s^{2\bullet}(J-1)/w_{i\Sigma\Sigma} \right) \right]$$

$$= (b^{(2)} + a)\left(1 - \frac{1}{I} \sum_i \sum_j \left(\frac{w_{ij\Sigma}}{w_{i\Sigma\Sigma}} \right)^2 \right), \tag{28}$$

$$E\left[\frac{1}{J} \sum_j \left(\sum_i \frac{w_{ij\Sigma}}{w_{\Sigma j\Sigma}}(y_{ijw} - y_{wjw})^2 - s^{2\bullet}(I-1)/w_{\Sigma j\Sigma} \right) \right]$$

$$= (b^{(1)} + a)\left(1 - \frac{1}{J} \sum_j \sum_i \left(\frac{w_{ij\Sigma}}{w_{\Sigma j\Sigma}} \right)^2 \right), \tag{29}$$

$$E\left[\sum_i \sum_j \frac{w_{ij\Sigma}}{w_{\Sigma\Sigma\Sigma}}(y_{ijw} - y_{www})^2 - s^{2\bullet}(IJ-1)/w_{\Sigma\Sigma\Sigma}\right]$$

$$= b^{(1)}\left(1 - \sum_i \left(\frac{w_{i\Sigma\Sigma}}{w_{\Sigma\Sigma\Sigma}}\right)^2\right)$$

$$+ b^{(2)}\left(1 - \sum_j \left(\frac{w_{\Sigma j\Sigma}}{w_{\Sigma\Sigma\Sigma}}\right)^2\right)$$

$$+ a\left(1 - \sum_i \sum_j \left(\frac{w_{ij\Sigma}}{w_{\Sigma\Sigma\Sigma}}\right)\right), \quad (30)$$

where $y_{iww} = \sum_j (w_{ij\Sigma}/w_{i\Sigma\Sigma})y_{ijw}$ and $y_{wjw} = \sum_i (w_{ij\Sigma}/w_{\Sigma j\Sigma})y_{ijw}$. To find the "unbiased estimator" of a, $b^{(1)}$ and $b^{(2)}$, we can drop the expectation operation of the above linear equations. As we can see, Dannenburg's estimates are based on the method of moments.

5.2. Empirical studies

Since Dannenburg's crossed classification model is of the form of linear mixed models, we could make use of the statistical packages that are designed especially for the parameter estimation in linear mixed models. One possibility is SAS. In our simulation studies, the results are obtained from the SAS procedure PROC MIXED. Since the simulation results for the ML and REML estimators are very similar, we only present the results for ML in this paper.

The estimation approaches we consider here are about the same as in Hachemeister's model, except that the first approach is Dannenburg's estimation approach. We would also provide two studies which is similar to Section 4. In Study 1, the error terms are simulated from multivariate normal distribution. In Study 2, the error terms are simulated from multivariate lognormal distribution.

5.2.1. Study 1

The simulation study is based on the following choice of parameters:

$$I = 12, \quad J = 8, \quad T_{ij} = n = 10,$$

$$b^{(1)} = 100, \quad b^{(2)} = 64, \quad a = 4, \quad s^2 = 196.$$

In this study, the observations are divided into $I \times J$ cells (96 cells). We randomly select 32 cells first, and these 32 cells have weight $w_{ijt} = 150$; then we select another 32 cells from the rest cells, these 32 cells have weight $w_{ijt} = 10$; the cells left have weight $w_{ijt} = 1.5$. Each sector retains its weight which has been assigned during the first replicate. The error terms ϵ_{ijt}s are simulated from multivariate normal distribution. The error structure is independent for Table 5, and exchangeable with $\rho = 0.4$ for Table 6.

As for the simulation results, we show the bias and mean square error (MSE) of the Dannenburg and ML approaches for β, $y_{ij,T_{ij}+1}$, z_{ij}, $z_i^{(1)}$, $z_j^{(2)}$, $b^{(1)}$, $b^{(2)}$, a and s^2.

We can see from Tables 5 and 6 that a significant advantage has been recorded for the ML approach over Dannenburg's approach. With regards to the structural parameters, the ML estimators have largely improved the estimation efficiency, especially for the parameters a and $b^{(2)}$. As a result, the performance of estimating the credibility factors and $y_{ij,T_{ij}+1}$ of the ML approach are very impressive. The reason for the poor performance of the Dannenburg estimator is that the level of precision in estimating a and $b^{(2)}$ is not enough to produce satisfactory estimates for the credibility factors. From our simulation results, for 500 repetitions, around 40% of the estimates of a are found to be negative, and around 6% of the estimates of $b^{(2)}$ are found to be negative. In contrast, all structural parameters estimated using ML approach fall in an admissible range.

From Table 5, we can see that MSE for a in Dannenburg's approach is about 500 times higher than the counterpart in the ML approach. As expected, the ML approach outperforms Dannenburg's estimation approach in estimating the future exposure y (relative efficiency around 1000) and the credibility factors (relative efficiency beyond 5000 for z_{ij} and $z_j^{(2)}$, relative efficiency beyond 40 for $z_i^{(1)}$). From Table 6, due to the correct assumption made on the error struc-

Table 5. Estimation results for Study 1 in the Dannenburg's model associated with an independent error structure and the observations are simulated from normal distribution

Parameter		Method		
		ML-I	ML-EX	Dannenburg
β	Bias	-1.42×10^{-1}	-1.54×10^{-1}	-2.23×10^{-1}
	MSE	1.55×10^{1} (1.48)	1.55×10^{1} (1.48)	2.30×10^{1}
y	Bias	-1.16×10^{-2}	-8.91×10^{-3}	-3.89
	MSE	3.96×10^{1} (960)	4.38×10^{1} (868)	3.80×10^{4}
z_{ij}	Bias	-1.80×10^{-3}	-2.08×10^{-3}	9.19×10^{-2}
	MSE	1.13×10^{-3} (> 5000)	1.30×10^{-3} (> 5000)	7.51×10^{1}
$z_{i}^{(1)}$	Bias	-2.34×10^{-3}	-2.35×10^{-3}	-5.54×10^{-3}
	MSE	3.54×10^{-5} (44.6)	3.68×10^{-5} (42.9)	1.58×10^{-3}
$z_{j}^{(2)}$	Bias	-4.32×10^{-3}	-4.37×10^{-3}	4.32×10^{-2}
	MSE	1.38×10^{-4} (> 5000)	1.42×10^{-4} (> 5000)	1.12
$b^{(1)}$	Bias	-5.88	-5.69	-1.23
	MSE	1.83×10^{3} (1.15)	1.82×10^{3} (1.15)	2.10×10^{3}
$b^{(2)}$	Bias	-4.42	-4.51	-1.26
	MSE	1.07×10^{3} (2.53)	1.07×10^{3} (2.53)	2.71×10^{3}
a	Bias	4.67×10^{-2}	5.56×10^{-2}	4.72×10^{-1}
	MSE	8.25×10^{-1} (506.67)	9.79×10^{-1} (426.97)	4.18×10^{2}
s^2	Bias	-1.62×10^{-1}	1.08×10^{-1}	1.65×10^{-1}
	MSE	8.80×10^{1} (1.03)	8.97×10^{1} (1.01)	9.03×10^{1}

Table 6. Estimation results for Study 1 in the Dannenburg's model associated with a MA(1) error structure ($\rho = 0.4$) and the observations are simulated from normal distribution

Parameter		Method		
		ML-I	ML-EX	Dannenburg
β	Bias	5.53×10^{-1}	5.43×10^{-1}	2.71×10^{-1}
	MSE	1.75×10^{1} (1.34)	1.76×10^{1} (1.33)	2.34×10^{1}
y	Bias	-2.48×10^{-1}	-2.44×10^{-1}	2.45×10^{1}
	MSE	3.52×10^{1} (> 10000)	3.78×10^{1} (> 10000)	1.21×10^{6}
z_{ij}	Bias	1.13×10^{-1}	4.29×10^{-2}	3.08×10^{-1}
	MSE	2.25×10^{-2} (203)	4.49×10^{-3} (> 1000)	4.57
$z_{i}^{(1)}$	Bias	-7.23×10^{-3}	-1.04×10^{-3}	-1.55×10^{-2}
	MSE	1.12×10^{-4} (286)	2.29×10^{-5} (> 1000)	3.20×10^{-2}
$z_{j}^{(2)}$	Bias	-8.76×10^{-3}	-1.78×10^{-3}	-5.65×10^{-2}
	MSE	1.89×10^{-4} (> 1000)	4.52×10^{-5} (> 10000)	8.37
$b^{(1)}$	Bias	-5.81	-5.96	-2.28
	MSE	1.75×10^{3} (1.14)	1.74×10^{3} (1.15)	2.00×10^{3}
$b^{(2)}$	Bias	-6.02	-6.18	-7.09
	MSE	9.16×10^{2} (2.70)	9.21×10^{2} (2.68)	2.47×10^{3}
a	Bias	3.36	-2.18×10^{-1}	4.20
	MSE	1.32×10^{1} (35.8)	1.05 (450)	4.72×10^{2}
s^2	Bias	-7.16×10^{1}	-7.41×10^{1}	-7.41×10^{1}
	MSE	5.13×10^{3} (1.07)	5.48×10^{3} (1.00)	5.48×10^{3}

ture, as we can expect that the ML-EX estimator performs the best. The ML-EX estimator maintains the high accuracy level in estimating the structural parameters, especially in estimating a. As a result, the MSE of the credibility factors in the ML-EX method is impressively low.

5.2.2. Study 2

In this study, the setting is similar to Study 1, except the error terms ϵ_{ijt}s are simulated from multivariate lognormal distribution. The vector of the error terms has mean shifted to **0**, and $s^2 = 196$. The error structure is independent in Table 3 and exchangeable with $\rho = 0.4$ in Table 4. The lognormal distribution has skewness of 2.97 and kurtosis of 25.3, which substantially departs from normal distribution. The estimators used in this study are the same as in Study 1.

As we explained in Study 1, Dannenburg's approach fails in providing credible estimates of a and $b^{(2)}$. From Table 7, we can observe the large discrepancies in the performance of the estimators for y and the credibility factors between the ML approach and Dannenburg's approach. The simulation shows even better results than we have observed in Table 5 in estimating the credibility factors (relative efficiency beyond 10,000 for z_{ij} and $z_j^{(2)}$, relative efficiency beyond 100 for $z_i^{(1)}$). From Table 8, the ML-EX outperforms the other methods especially in estimating a and $z_i^{(1)}$. Therefore, the simulation results reaffirm that the proposed ML approach can provide us credible estimates even when the distribution of the observation substantially deviates from normality.

6. Concluding remarks

In this paper, we implement the linear mixed model in credibility context, and use ML and REML approach to estimate the structural parameters. There are other approaches in estimating the structural parameters in the credibility models. By comparing our approaches with the

generalized least square estimation approach proposed by Cossette and Luong (2003) and the GEE approach proposed by Lo, Fung, and Zhu (2006) and Lo, Fung, and Zhu (2007), we demonstrate the merits of our approaches. The former can hardly be extended beyond the Bühlmann model, in which the heteroscedasticity is assumed in the error terms. The latter is hard to apply to classical credibility models when the number of observations with regard to the same contract gets bigger. For instance, if the number of observations for the same contract exceeds 10, the working covariance matrix would be extremely complicated, and the dimension would be very large in the GEE approach. Also the robustness of these two approaches has not been investigated.

Furthermore, from the empirical studies, the time our approach takes is much shorter than the GEE approach. For instance, it takes less than 15 minutes to get the ML and REML estimation results for 500 repetitions in the Hachemeister model using a Pentium 4 3.00 GHz desktop computer with 2.00 GB of RAM; however it takes more than one and a half hours to get the GEE estimates for 500 repetitions. Furthermore, with the aid of software, there are no additional complications when we want to exercise the proposed ML and REML approaches with different assumptions on the error structure.

Moreover, we have investigated the performance of ML and REML methods when the assumptions regarding the error structure and distribution are violated. We can see from the simulation studies, for the situations that the error terms follow normal and non-normal distributions, ML and REML methods maintain satisfactory results. This serves an empirical justification of using the ML and REML approaches when distribution of the observations is unknown.

In this paper, we have only showed the results of the ML approach for brevity. Verbeke and Molenberghs (2000) made the comparison

Table 7. Estimation results for Study 2 in the Dannenburg's model associated with an independent error structure and the observations are simulated from lognormal distribution

Parameter		Method		
		ML-I	ML-EX	Dannenburg
β	Bias	-2.19×10^{-1}	-2.20×10^{-1}	-2.67×10^{-1}
	MSE	1.57×10^1 (1.54)	1.57×10^1 (1.54)	2.41×10^1
y	Bias	2.31×10^{-2}	2.37×10^{-2}	-7.55
	MSE	4.07×10^1 (946)	4.07×10^1 (946)	3.85×10^4
z_{ij}	Bias	-2.44×10^{-3}	-3.40×10^{-3}	2.12×10^{-1}
	MSE	9.93×10^{-4} (>10000)	1.24×10^{-3} (>10000)	1.74×10^1
$z_i^{(1)}$	Bias	-2.16×10^{-3}	-2.13×10^{-3}	-3.13×10^{-3}
	MSE	4.28×10^{-5} (107)	4.24×10^{-5} (108)	4.57×10^{-3}
$z_j^{(2)}$	Bias	-3.80×10^{-3}	-3.76×10^{-3}	-7.04×10^{-2}
	MSE	1.22×10^{-4} (>10000)	1.22×10^{-4} (>10000)	4.47
$b^{(1)}$	Bias	-4.81	-4.84	-6.67×10^{-2}
	MSE	1.85×10^3 (1.12)	1.85×10^3 (1.12)	2.07×10^3
$b^{(2)}$	Bias	-3.56	-3.54	2.69
	MSE	9.96×10^2 (2.90)	9.96×10^2 (2.90)	2.89×10^3
a	Bias	-7.10×10^{-3}	-2.07×10^{-2}	1.96×10^{-1}
	MSE	6.61×10^{-1} (539)	7.95×10^{-1} (493)	3.92×10^2
s^2	Bias	-3.41×10^{-1}	-4.33×10^{-1}	-4.34×10^{-1}
	MSE	1.73×10^2 (0.97)	1.67×10^2 (1.00)	1.67×10^2

Table 8. Estimation results for Study 2 in the Dannenburg's model associated with a MA(1) error structure ($\rho = 0.4$) and the observations are simulated from lognormal distribution

Parameter		Method		
		ML-I	ML-EX	Dannenburg
β	Bias	2.64×10^{-1}	2.48×10^{-1}	3.29×10^{-1}
	MSE	1.70×10^1 (1.44)	1.70×10^1 (1.44)	2.45×10^1
y	Bias	7.04×10^{-3}	-2.13×10^{-3}	2.69
	MSE	2.57×10^1 (>1000)	2.91×10^1 (>1000)	9.15×10^4
z_{ij}	Bias	1.68×10^{-1}	5.31×10^{-2}	-7.71×10^{-1}
	MSE	5.26×10^{-2} (>10000)	6.78×10^{-3} (>100000)	1.96×10^3
$z_i^{(1)}$	Bias	-1.69×10^{-2}	-1.27×10^{-3}	-1.81×10^{-2}
	MSE	4.93×10^{-4} (3.23)	2.19×10^{-5} (72.60)	1.59×10^{-3}
$z_j^{(2)}$	Bias	-2.02×10^{-2}	-2.41×10^{-3}	3.33×10^{-2}
	MSE	9.48×10^{-4} (>1000)	1.06×10^{-4} (>10000)	4.37
$b^{(1)}$	Bias	-4.53	-5.05	-2.22
	MSE	1.69×10^3 (1.12)	1.66×10^3 (1.14)	1.90×10^3
$b^{(2)}$	Bias	-3.06	-3.54	-8.13
	MSE	1.01×10^3 (2.96)	1.01×10^3 (2.96)	2.99×10^3
a	Bias	9.74	7.10×10^{-2}	9.19
	MSE	1.22×10^2 (4.43)	1.51 (358)	5.40×10^2
s^2	Bias	-7.59×10^1	-7.75×10^1	-7.75×10^1
	MSE	5.84×10^3 (1.04)	6.09×10^3 (1.00)	6.09×10^3

between ML and REML estimation. With regard to the mean squared error of estimating the variance and covariance parameters, neither of the two estimation procedures are universally better than the other. The performance of ML and REML depends on the specification of the underlying model, and possibly on the true value of the variance and covariance parameters. However, when the rank of design matrix $\mathbf{X}_i$ is less than 4, the ML estimator of the residual σ^2 generally outperforms the REML estimator, but the opposite is true when the rank of $\mathbf{X}_i$ gets larger. Generally speaking, we can expect the difference between ML and REML estimator increases when the rank of $\mathbf{X}_i$ increases. In our simulation studies, since the rank of $\mathbf{X}_i$ is not large, neither the ML approach or the REML approach performs universally better than the other in both Dannenburg's model and Hachemeister model.

References

Antonio, K., and J. Beirlant, "Actuarial Statistics with Generalized Linear Mixed Models," *Insurance: Mathematics and Economics* 40, 2006, pp. 58–76.

Cossette, H., and A. Luong, "Generalized Least Squares Estimators for Covariance Parameters for Credibility Regression Models with Moving Average Errors," *Insurance: Mathematics and Economics* 32, 2003, pp. 281–293.

Dannenburg, D. R., R. Kaas, and M. J. Goovaerts, *Practical Actuarial Credibility Models*, Institute of Actuarial Science and Econometrics, University of Amsterdam, 1996.

Frees, E. W., V. R. Young, and Y. Lou, "A Longitudinal Data Analysis Interpretation of Credibility Models," *Insurance: Mathematics and Economics* 24, 1999, pp. 229–247.

Hachemeister, C. A., "Credibility for Regression Models with Application to Trend," In *Credibility: Theory and Applications*, P. M. Kahn (Ed.), New York: Academic, 1975, pp. 129–163.

Laird, N. M., and J. H. Ware, "Random-Effect Models for Longitudinal Data" *Biometrics* 38, 1982, pp. 963–974.

Lindstrom, M. J., and D. M. Bates, "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *Journal of the American Statistical Association* 83, 1988, pp. 1014–1022.

Lo, C. H., W. K. Fung, and Z. Y. Zhu, "Generalized Estimating Equations for Variance and Covariance Parameters in Credibility Models," *Insurance: Mathematics and Economics* 39, 2006, pp. 99–113.

Lo, C. H., W. K. Fung, and Z. Y. Zhu, "Structural Parameter Estimation Using Generalized Estimating Equations for Regression Credibility Models," *ASTIN Bulletin* 37, 2007, pp. 323–343.

McCulloch, C. E., "Maximum Likelihood Algorithms for Generalized Linear Mixed Models," *Journal of the American Statistical Association* 92, 1997, pp. 162–170.

Nelder, J. A., and R. Mead, "A Simplex Algorithm for Function Minimization," *Computer Journal* 7, 1965, pp. 308–313.

Rao, C. R., "Simultaneous Estimation of Parameters in Different Linear Models and Applications to Biometric Problems," *Biometrics* 31, 1975, pp. 545–554.

Verbeke, G., and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, New York: Springer, 2000.